A simple approach for detecting HLA-A02 alleles in archival formalin-fixed paraffin-embedded tissue samples and an application example for studying cancer immunoediting

Witt, Johannes

2023-01

Witt , J , Haupt , S , Ahadova , A , Bohaumilitzky , L , Fuchs , V , Ballhausen , A , Przybilla , M J , Jendrusch , M , Seppälä , T T , Fuerst , D , Walle , T , Busch , E , Haag , G M , Hueneburg , R , Nattermann , J , Doeberitz , M V K , Heuveline , V & Kloor , M 2023 , ' A simple approach for detecting HLA-A02 alleles in archival formalin-fixed paraffin-embedded tissue samples and an application example for studying cancer immunoediting ' , HLA , vol. 101 , no. 1 , pp. 24-33 . https://doi.org/10.1111/tan.14846

http://hdl.handle.net/10138/352487
https://doi.org/10.1111/tan.14846

Received: 26 May 2022 | Revised: 5 September 2022 | Accepted: 7 October 2022

DOI: 10.1111/tan.14846

ORIGINAL ARTICLE

HLA Immune Response Genetics WILEY

A simple approach for detecting *HLA-A**02 alleles in archival formalin-fixed paraffin-embedded tissue samples and an application example for studying cancer immunoediting

Johannes Witt[1,2] | **Saskia Haupt**[3,4] | **Aysel Ahadova**[1,2] | **Lena Bohaumilitzky**[1,2] | **Vera Fuchs**[1,2] | **Alexej Ballhausen**[1,2,5] | **Moritz Jakob Przybilla**[1,2] | **Michael Jendrusch**[1,2] | **Toni T. Seppälä**[6,7,8,9] | **Daniel Fürst**[10] | **Thomas Walle**[11] | **Elena Busch**[11] | **Georg Martin Haag**[11,12] | **Robert Hüneburg**[13,14] | **Jacob Nattermann**[13,14] | **Magnus von Knebel Doeberitz**[1,2] | **Vincent Heuveline**[3,4] | **Matthias Kloor**[1,2]

[1]Department of Applied Tumor Biology, Institute of Pathology, Heidelberg University Hospital, Heidelberg, Germany

[2]Clinical Cooperation Unit Applied Tumor Biology, German Cancer Research Center (DKFZ), Heidelberg, Germany

[3]Engineering Mathematics and Computing Lab, Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University, Heidelberg, Germany

[4]Data Mining and Uncertainty Quantification (DMQ), Heidelberg Institute for Theoretical Studies, Heidelberg, Germany

[5]Department of Hematology, Oncology and Tumor Immunology, Charité -Universitätsmedizin Berlin, Berlin, Germany

[6]Faculty of Medicine and Health Technology, University of Tampere, Tampere, Finland

[7]Department of Gastroenterology and Alimentary Tract Surgery, Tampere University Hospital, Tampere, Finland

[8]Department of Gastrointestinal Surgery, Helsinki University Central Hospital, Helsinki, Finland

[9]Applied Tumor Genomics Research Program, University of Helsinki, Helsinki, Finland

[10]Department of Transplantation Immunology, Institute of clinical Transfusion Medicine and Immune Genetics, Ulm, Germany

[11]Department of Medical Oncology, National Centre for Tumor Diseases, Heidelberg University Hospital, Heidelberg, Germany

[12]Clinical Cooperation Unit Applied Tumor-Immunity, German Cancer Research Center (DKFZ), Heidelberg, Germany

[13]Department of Internal Medicine I, University Hospital Bonn, Bonn, Germany

[14]National Center for Hereditary Tumor Syndromes, University Hospital Bonn, Bonn, Germany

Correspondence
Johannes Witt, Department of Applied Tumor Biology, Institute of Pathology, Heidelberg University Hospital, Heidelberg, Germany.
Email: johannes.witt@med.uni-heidelberg.de

The HLA system represents a central component of the antigen presentation machinery. As every patient possesses a defined set of HLA molecules, only certain antigens can be presented on the cell surface. Thus, studying HLA type-dependent antigen presentation can improve the understanding of variation in susceptibility to various diseases, including infectious diseases and

Abbreviations: cMS, coding microsatellite; CRC, colorectal cancer; FFPE, formalin-fixed paraffin-embedded; FN, false negative; FP, false positive; FSP, frameshift peptide; MSI, microsatellite instability; SBT, sequencing-based typing; SNP, single nucleotide polymorphism.

Funding information
Else-Kröner-Fresenius Foundation, Grant/Award Number: 2018_A44; German Cancer Aid, Grant/Award Number: 70113455; Klaus Tschira Foundation, Grant/Award Number: 00.012.2021; Sigrid Juselius Foundation, Jane and Aatos Erkko Foundation, Finnish Cancer Society, Relander Foundation, Emil Aaltonen Foundation, Finnish Medical Foundation, State Research Funding, Grant/Award Number: TYH2022323; Academy of Finland, Grant/Award Number: 33867

cancer. In archival formalin-fixed paraffin-embedded (FFPE) tissue, the HLA type is difficult to analyze because of fragmentation of DNA, hindering the application of commonly used assays that rely on long DNA stretches. Addressing these difficulties, we present a refined approach for characterizing presence or absence of *HLA-A**02, the most common HLA-A allele in the Caucasian population, in archival samples. We validated our genotyping strategy in a cohort of 90 samples with HLA status obtained by an NGS-based method. 90% (n = 81) of the samples could be analyzed with the approach. For all of them, the presence or absence of *HLA-A**02 alleles was correctly determined with the method, demonstrating 100% sensitivity and specificity (95% CI: 91.40%–100% and 91.19%–100%). Furthermore, we provide an example of application in an independent cohort of 73 FFPE microsatellite-unstable (MSI) colorectal cancer samples. As MSI cancer cells encompass a high number of mutations in coding microsatellites, leading to the generation of highly immunogenic frameshift peptide antigens, they are ideally suited for studying relations between the mutational landscape of tumor cells and interindividual differences in the immune system, including the HLA genotype. Overall, our method can help to promote studying HLA type-dependency during the pathogenesis of a wide range of diseases, making archival and historic tissue samples accessible for identifying *HLA-A**02 alleles.

KEYWORDS
cancer immunoediting, formalin-fixed paraffin-embedded tissue samples, HLA typing, MSI cancer

1 | INTRODUCTION

An essential component of the immune system is the HLA system, which is responsible for the presentation of self- and nonself-antigens on the cell surface.[1] Determining the HLA type is of central interest both in clinical fields, such as solid organ transplantation, but also in disease association research.[2] Apart from autoimmune diseases, investigating the HLA type as a possible risk modulator and prognostic factor has gained attention particularly in cancer research.[3–6] As routinely applied DNA typing methods are not applicable when using DNA derived from formalin-fixed paraffin-embedded (FFPE) material,[7] methods for identifying HLA alleles in this type of tissue material could be of high value.

In principle, every individual has a specific set of HLA class I molecules encoded by the paternal and maternal alleles of the highly polymorphic genes *HLA-A*, *HLA-B* and *HLA-C*. HLA molecules can bind only a limited number of peptides, called the peptide repertoire. It is mainly determined by the molecular structure of the peptide-binding groove.[8] Accordingly, only a certain type of antigens can be presented on the cell surface to the host's immune cells,[1,8] thereby possibly influencing the susceptibility of an individual to certain diseases. HLA class I molecules with a strongly overlapping peptide repertoire constitute a supertype.[9,10] For HLA-A molecules, there are four major supertypes (A01, A02, A03 and A24).

The important role of the HLA type in determining the clinical course of infectious diseases has been investigated in many entities, such as hepatitis B or HIV.[11] However, by determining the repertoire of tumor antigens presented on the cell surface, HLA genotype might also influence the course of cancer progression and the mutational spectrum of arising tumor cell clones during carcinogenetic evolution. Recent studies reported an HLA type-dependent negative selection of oncogenic mutations, predominantly examining missense mutations.[12–14] In clinical cancer research, investigating HLA type as a predictive marker for the response to immune checkpoint inhibitors has recently gained attention.[15] In addition, we previously found biological evidence for counterselection of highly immunogenic frameshift mutations in microsatellite-unstable (MSI) cancer.[16] MSI tumors arise because of inactivation of the DNA mismatch repair system leading to a

high number of somatic mutations, especially located at microsatellites.[17,18] Mutations in coding microsatellites (cMS) promote malignant transformation, but also increase the visibility of cancer cells to the immune system.[19] This is because of the generation of novel peptides, resulting from a translational frameshift (frameshift peptides, FSPs).[19–22] Because of this distinct pathogenesis, MSI tumors are highly immunogenic and represent an ideal model for studying immunoediting and possible immune escape mechanisms, such as mutations in the *B2M* gene or downregulation of the HLA class I heavy chain.[23–25]

DNA derived from FFPE samples is of limited quality, mainly because of DNA fragmentation resulting from the formalin fixation process.[26] Given the possibility to revisit large sample collections of archival FFPE cancer samples, methods that overcome this hurdle would allow studying the implications of the HLA genotype during historic disease outbreaks, but also tumor evolution across cancer types. Previously, a Sanger sequencing-based approach for determining the HLA-A alleles of FFPE probes was presented.[7] To achieve a greater practicability by reducing the number of required reactions and preventing the possibility of ambiguous typing results, we here present a simple and highly accurate method for evaluating the presence of *HLA-A**02 alleles in FFPE tissue probes. We focused on *HLA-A**02 for three reasons: First, in MSI CRC, particularly downregulation of HLA-A molecules has been shown to be a common immune escape mechanism.[25,27] This potentially indicates a pronounced selective pressure especially for this HLA class I locus. Second, *HLA-A**02 appears to be of prognostic relevance in different malignancies, such as in advanced ovarian carcinoma[28] and prostate cancer.[29] Finally, being the most common HLA class I allele with a homo- and heterozygote frequency of approximately 48.6% in a representative German population,[30] determining *HLA-A**02 allows to divide a cohort of patients into two groups of comparable size.

The approach for identifying *HLA-A**02 alleles is based on determining one recently described single nucleotide polymorphism (SNP).[31] In addition, we show an exemplary application of the method in the context of tumor immunoediting by analyzing the influence of the *HLA-A**02 status on cMS mutation frequency in MSI colorectal cancer (CRC) samples.

2 | METHODS

Tumor specimens. Seventy-three MSI CRC samples (primary site) with previously published somatic mutation data were reanalyzed in this study (https://github.com/atb-data/neoantigen-landscape-msi).[16] Thirty-one of the CRC cases are Lynch syndrome-associated, whereas 41 occurred sporadically (no information for one patient, clinical data in Tables S2a and S2b). The study was approved by the institutional Ethics Committee. Written informed consent was obtained from all patients.

Analysis of HLA-A allele sequences. Aligned sequences of all HLA-A alleles were obtained from the IPD-IMGT/HLA database (release 3.49, downloaded on August 14th, 2022)[32] and analyzed with Biopython (version 1.76)[33] regarding mismatches in the primer binding site of A2-2F or G-R and the SNP described earlier.[31]

HLA-A**02 status. The presence of at least one *HLA-A**02 allele was determined by analyzing one previously described SNP at the 5′-end of exon 2 of the *HLA-A* gene.[31] Initial PCR with an amplicon length of 134 bp was performed with the forward primer A2-2F 5′-TCTCAGCCACTCCTCGTC-3′[7] (primer binding site in intron 1 of the *HLA-A* gene) and the reverse primer G-R 5′-TGTCGAACCGCACGAACTG-3′ (primer binding site in exon 2 of the *HLA-A* gene, for more details see supplementary material). G-R was designed using Geneious prime® (version 2020.0.5). Sanger sequencing was only performed with the reverse primer G-R on an ABI3130*xl* genetic analyzer (Applied Biosystems, Darmstadt, Germany). Thus, the obtained sequencing chromatogram corresponds to the reverse-complement of the coding strand (see Figure S1). Samples were assigned to the two groups G_{pr} and G_{non} according to the nucleotide at position 78 of the *HLA-A* gene (reference sequence NM_001242758.1, see Figure S1): samples having at least one *HLA-A**02 allele are assigned to the group G_{pr} (pr = present, homo-/heterozygous for "T" at pos. 78, "A" in the reverse-complement) and samples without any *HLA-A**02 allele belong to the group G_{non} (non = non-present, no "T" at pos. 78, no "A" in the reverse-complement). Samples with low Sanger sequencing quality or background interference were excluded (for details see Figure S1). The sequencing-based typing approach was experimentally validated in a blinded setting by analyzing 90 samples of known HLA type, determined from blood samples by an NGS-based method routinely used in the clinical setting.

Average frequency of m1-frameshift mutations. We focused on cMS mutations with the translational reading frame "minus 1" (m1, deletion of 1 nucleotide or insertions of two nucleotides), which are considered the predominant type of frameshift mutations in MSI cancer.[16] For each of the 41 cMS analyzed in Ballhausen et al.,[16] the average frequency of m1-mutations $\overline{f}_{m1,G}$, leading to an FSP m, was determined. $\overline{f}_{m1,G}$ was calculated for all patients (G_{all}), and samples of the G_{pr} and the G_{non} group (formula in supplementary material).

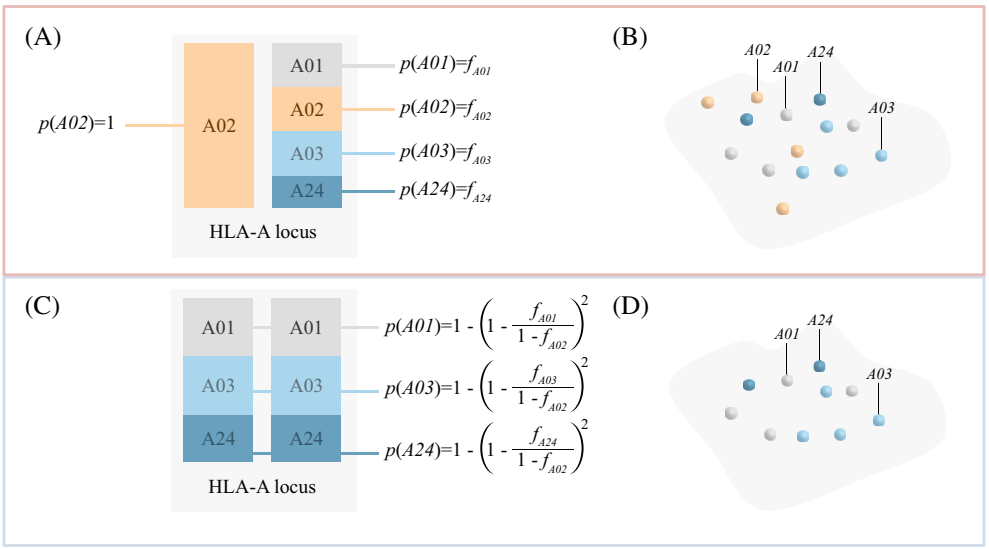

FIGURE 1 Possible HLA-A supertype constellations in patients with at least one *HLA-A**02 allele (G_{pr} group, red box) and in patients without any *HLA-A**02 allele (G_{non} group, blue box). (A) and (C) show the HLA-A locus, which encompasses one maternal and one paternal allele, in the G_{pr} (A) and the G_{non} (C) group. Patients of the G_{pr} group have at least one allele of the supertype A02. As the second allele remains unclear, it can belong to the HLA-A supertypes A01, A02, A03 or A24. By contrast, patients of the G_{non} group do not have any A02 allele, illustrated by the absence of an orange bar in (C). Thus, patients of this group only possess alleles of the HLA-A supertypes A01, A03 or A24. The probability $p(S)$ that a supertype S is present in a patient of the corresponding group is shown for each group separately (details in supplementary material). (B) and (D) illustrate the distribution of supertypes in both subclasses. Patients of the G_{non} group do not possess any A02 allele (illustrated by the absence of orange balls in (D)). Thus, the likelihood to have an allele of the supertypes A01, A03 and A24 is higher in this group compared with the general population.

Selection of cMS. All 41 cMS in 40 target genes from Ballhausen et al.[16] were analyzed. In addition, the candidates with an average m1-mutation frequency $\overline{f}_{m1,G_{all}} > 0.25$ were examined separately. The corresponding target genes are *ACVR2A*, *TGFBR2*, *SLC35F5*, *ASTE1*, *CEP164*, *AIM2*, *MARCKS*, *SLC22A9*.

MHC ligand prediction data. Regarding the 41 m1-FSPs, binding affinity prediction data for the four supertype representatives *HLA-A**01:01 (A01), *HLA-A**02:01 (A02), *HLA-A**03:01 (A03), *HLA-A**24:02 (A24) were obtained from NetMHCpan-4.1,[34] a state-of-the-art in silico tool for simulating binding between peptides and HLA molecules, as described previously.[16] The corresponding FSPs including 8 wildtype amino acids at the *N*-terminus are shown in Table S3. The column "Score_EL" (here called epitope likelihood EL_S) was extracted from the output XLS file,[34] estimating the likelihood that a peptide is a ligand for the representative of the supertype *S*.

Immunological scores for the HLA-A locus. We defined scores for estimating the immunogenicity of the examined FSPs in patients of the G_{pr} and in patients of the G_{non} group according to the particular HLA-A constellation (see Figure 1). Considering the group-specific frequency of HLA-A alleles, we determined the overall ligand likelihood $OLL_G(m)$. It estimates the probability that at least one peptide derived from the FSP m is presented by an HLA-A allele in a patient belonging to the group $G \in \{G_{pr}, G_{non}\}$ (for details, see supplementary material).

3 | RESULTS

Overview of the method for determining the *HLA-A*02* status. Song et al.[31] developed an approach based on sequence-specific primers for identifying samples with *HLA-A**02 alleles. However, when FFPE tissue is the only material available, a PCR setting with an amplicon length of 881/882 bp is not suitable because of DNA fragmentation, which hampers amplification of long DNA stretches. In order to address this issue, we designed a two-level sequencing-based typing approach (Figure 2): The initial PCR with the primers A2-2F and G-R aims at amplifying all *HLA-A**02 alleles, which is a precondition for achieving high sensitivity. As a second step, Sanger sequencing of the described SNP at position 78 of the *HLA-A* gene, located within the PCR fragment, allows to specifically distinguish *HLA-A**02 alleles from other alleles possibly amplified in the initial PCR. Samples with at least one *HLA-A**02 allele are assigned to the G_{pr} group, whereas samples without any *HLA-A**02 allele belong to the group G_{non}. The

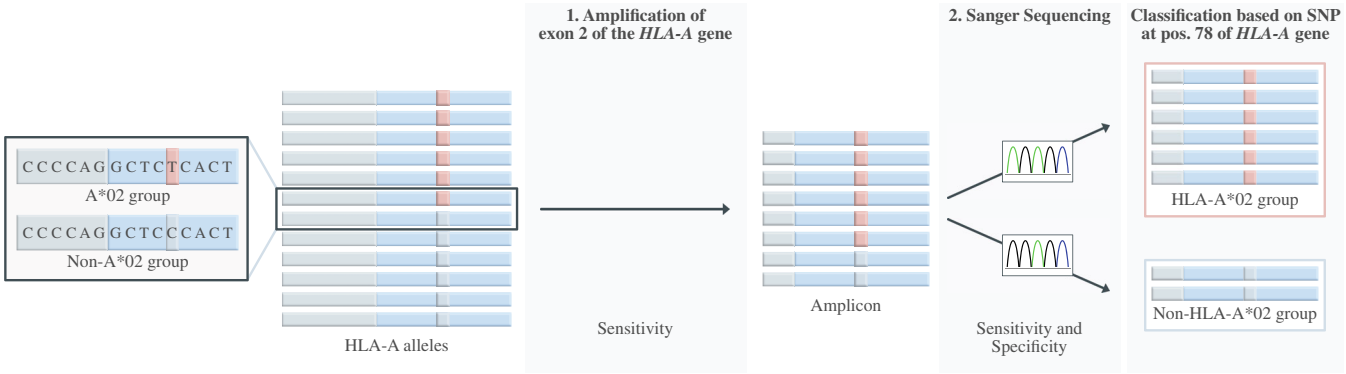

FIGURE 2 A two-step sequencing-based typing approach for reliable identification of samples with *HLA-A**02 alleles. *HLA-A**02 alleles possess a "T" at position 78 of the *HLA-A* gene (depicted in red), whereas Non-*HLA-A**02 alleles show another nucleotide (e.g., a "C", depicted in gray). The initial PCR A2-2F/G-R represents a precondition for achieving high sensitivity as all *HLA-A**02 alleles have to be amplified. Also Non-*HLA-A**02 alleles might generate an amplicon. Thus, high specificity (and sensitivity) is ensured by subsequent sequencing and classification according to the SNP at position 78 of the *HLA-A* gene.

resulting sequencing chromatogram with the SNP is shown in Figure S1.

As an advantage compared with an approach based on sequence-specific primers, reliable annealing of primers, which is influenced by PCR conditions, is not a central prerequisite for achieving high specificity. This means that the assignment of samples to the G_{non} group is not based on the absence of an amplicon in the initial PCR. Conversely, unspecific binding of primers cannot lead to false positive (FP) classification results (patients with two Non-*HLA-A**02 alleles assigned to the G_{pr} group). Accordingly, it is possible to choose a relatively low, and thus permissive annealing temperature, potentially amplifying HLA-A alleles with mismatches in the primer binding site of A2-2F or G-R.

Evaluation and experimental validation of the method. Because of the plethora of HLA-A alleles, an approach for identifying a certain allele group should be evaluated concerning possible candidates that would lead to FP or false negative (FN) classification results. For this purpose, we systematically analyzed aligned HLA-A allele sequences, published in the IPD-IMGT/HLA database.[32] Generally, nucleotide changes arising from formalin fixation are unlikely to hinder the correct identification of *HLA-A**02 alleles because of the abundance of unchanged templates[26] (see FN (1) in Figure S3 and FP (1) in Figure S4). Nevertheless, two scenarios possibly result in the assignment of *HLA-A**02 alleles to the G_{non} group, here referred to as FN classification results: First, *HLA-A**02 alleles that are not amplified in the PCR A2-2F/G-R would not be recognized in the sequencing step (FN (2) in Figure S3). However, 27 from the observed 28 mismatches between A2-2F or G-R and their corresponding primer binding site are unlikely to hamper amplification[35,36] (mismatches between *HLA-A**02 alleles and the primers are highlighted in Tables S6 and S7). Only the A2-2F primer binding site of *HLA-A**02:01:207 contains two mismatches, which is likely to prevent PCR product formation.[35] Moreover, 11 (1.2%) of the 955 *HLA-A**02 alleles examined have a base other than "T" at position 78 of the *HLA-A* gene, resulting in the assignment of these alleles to the G_{non} group (FN (3) in Figure S3, see Table S8). Conversely, 6 (0.2%) of the 3343 Non-*HLA-A**02 alleles have a "T" at position 78 of the *HLA-A* gene, potentially entailing FP classification results (*HLA-A**01:01:13, -*A**11:199:03, -*A**30:08:01, -*A**30:166, -*A**32:01:43, -*A**68:03:02, FP (2) in Figure S4). Three of these have an exactly matching primer-binding site with A2-2F and G-R (*HLA-A**01:01:13, -*A**30:166, -*A**68:03:02).

The sequencing-based typing approach for determining the *HLA-A**02 status of FFPE tissue samples was experimentally validated in a cohort of 90 patients with known HLA-A type. The latter was obtained by a routinely applied NGS-based approach using DNA from blood samples. From these 90 samples, 81 were technically analyzable (90%). All of them were assigned to the correct group G_{pr} (n = 41, 100% sensitivity, 95% CI: 91.40%–100%) or G_{non} (n = 40, 100% specificity, 95% CI: 91.19%–100%; see Table S1). For the remaining samples, PCR amplification failed (n = 5, 5.6%) or resulted in background noise in the sequencing chromatogram preventing reliable classification (n = 4, 4.4%).

Example of application in tumor biology research: Analyzing *HLA-A02-dependent negative selection in MSI CRC**. After validating the accuracy of the developed classification approach, we aimed to demonstrate its applicability in current molecular biology research. A central controversy in the context of tumor immunology is whether the individual's HLA type influences the mutational

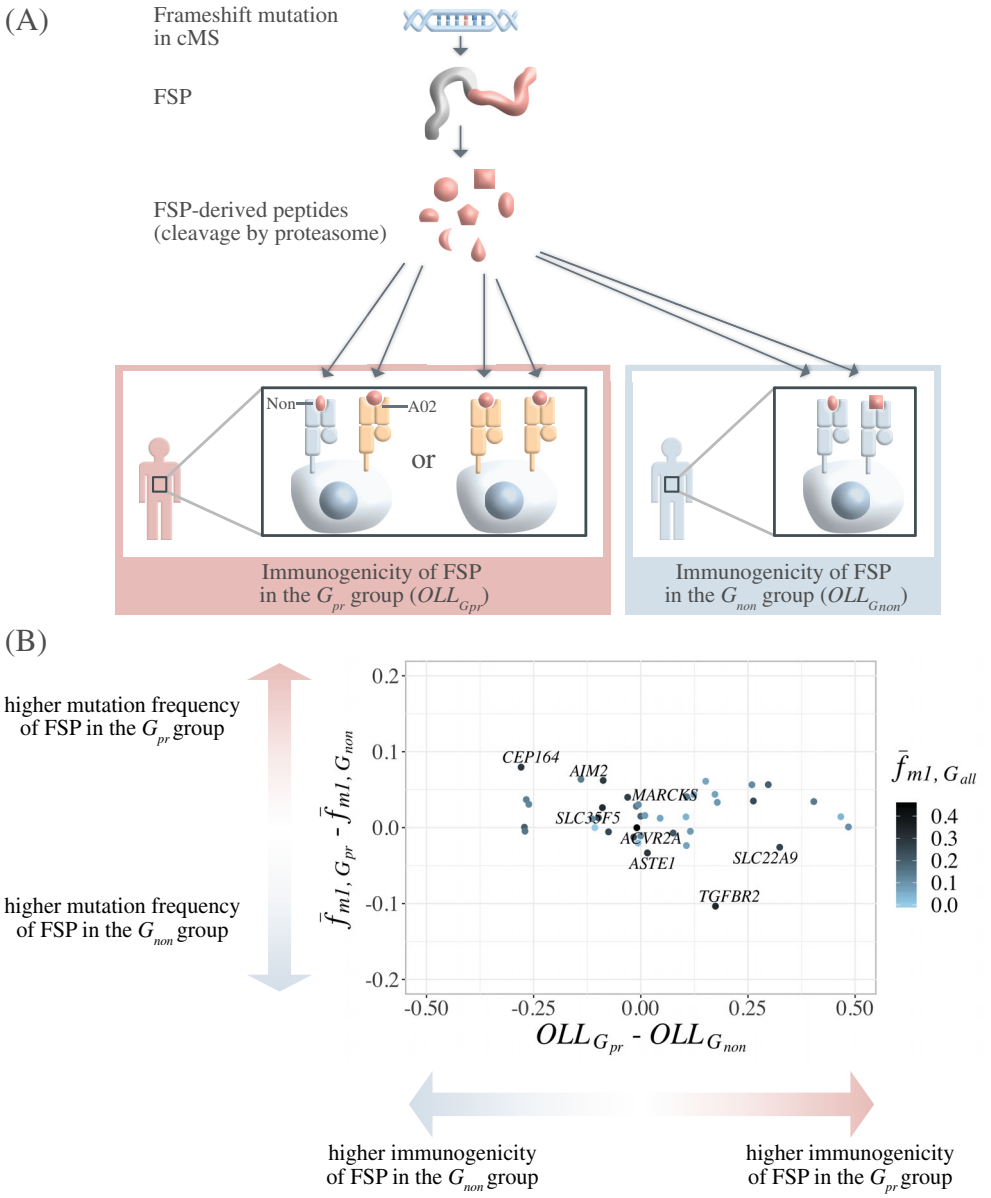

FIGURE 3 (A) Estimating the immunogenicity of FSPs in patients of the G_{pr} and the G_{non} group with the overall ligand likelihood OLL_G. Frameshift mutations in cMS lead to the generation of novel FSP antigens. Patients of the G_{pr} group (red box) have at least one A02 allele (depicted in orange), whereas patients of the G_{non} group (blue box) only have two Non-A02 alleles (depicted in blue). The immunogenicity of a given FSP in each patient group was estimated by checking the compatibility of all possible FSP-derived peptides (8–14 amino acids, illustrated by different red forms) with the HLA-A molecules given in one group. (B) Association between the absolute difference of the estimated immunogenicity of FSPs (x-axis, $\text{OLL}_{G_{\text{pr}}} - \text{OLL}_{G_{\text{non}}}$) and the absolute difference of the underlying cMS mutation frequency (y-axis, $\bar{f}_{m1,G_{\text{pr}}} - \bar{f}_{m1,G_{\text{non}}}$) between both groups G_{pr} and G_{non}. All samples are included. The average frequency of m1-mutations $\bar{f}_{m1,G_{\text{all}}}$, irrespective of *HLA-A**02 status, corresponds to the intensity of the data points (see legend on the right). The candidates with an m1-mutation frequency higher than 0.25, which were analyzed separately, are labeled. There is no correlation between $\text{OLL}_{G_{\text{pr}}} - \text{OLL}_{G_{\text{non}}}$ and $\bar{f}_{m1,G_{\text{pr}}} - \bar{f}_{m1,G_{\text{non}}}$, considering all 41 candidates ($p = 0.4985$, Pearson's $r = -0.0007$). However, when selecting the cMS with an average frequency of m1-mutations higher than 0.25, one can see an inverse correlation ($p = 0.01231$, Pearson's $r = -0.77254$).

landscape of cancers.[12–14,37] With the developed approach, we determined the *HLA-A**02 status of an independent cohort of 73 MSI CRC samples. Thirty-six of these had at least one *HLA-A**02 allele (group G_{pr}, 12 hereditary and 24 sporadic cases) and 37 samples did not possess any *HLA-A**02 allele (group G_{non}, 19 hereditary and 17 sporadic cases, one patient without information on etiology of MSI, see Table S2a). The proportion of hereditary MSI cases did not deviate significantly between both patient groups G_{pr} and G_{non} ($p = 0.1528$ according to two-sided Fisher's test).

We compared the frequency of cMS mutations and the predicted immunogenicity of the resulting FSPs

between the G_{pr} and the G_{non} patient group. For approximating the immunogenicity of the 41 FSPs, we modified recently published scores[16] and defined the overall ligand likelihood OLL_G to account for the particular HLA-A constellation in both patient groups G_{pr} and G_{non} (see Methods and Table S4). It estimates the group-specific likelihood that at least one cleaved peptide derived from a certain FSP is presented by the HLA-A molecules of patients in the G_{pr} or in the G_{non} group (see Figure 3A). Thus, the score not only considers epitope predictions for alleles of the HLA-A supertype A02, but also for the other HLA-A supertypes according to their probability to be present in patients of the G_{pr} or the G_{non} group (described in Figure 1 in the methods).

Next, we compared the frequency of cMS mutations and their HLA-A-dependent immunogenicity in the G_{pr} and G_{non} group. For this purpose, we calculated the differences between the overall ligand likelihood OLL_G in both groups ($OLL_{G_{pr}} - OLL_{G_{non}}$) and between the average frequency of m1-mutations $\overline{f}_{m1,G}$ in both groups ($\overline{f}_{m1,G_{pr}} - \overline{f}_{m1,G_{non}}$). High values for $OLL_{G_{pr}} - OLL_{G_{non}}$ indicate that a given FSP is better visible to the immune system in patients of the G_{pr} than in the G_{non} group (illustrated by the arrows next to the *x*-axis in Figure 3B). Similarly, high values for $\overline{f}_{m1,G_{pr}} - \overline{f}_{m1,G_{non}}$ mean that the underlying cMS mutation occurs more frequently in the G_{pr} group, and vice versa (illustrated by the arrows next to the *y*-axis in Figure 3B). Considering all 41 cMS, no correlation is detectable between $OLL_{G_{pr}} - OLL_{G_{non}}$ and $\overline{f}_{m1,G_{pr}} - \overline{f}_{m1,G_{non}}$. This applies both to all patients of the cohort ($p = 0.4985$, Pearson's $r = -0.0007$, see Figure 3B) and to only *B2M*-mutated samples ($p = 0.3060$, Pearson's $r = -0.0814$, see Figure S5). However, as the average frequency of m1-mutations $\overline{f}_{m1,G_{all}}$, irrespective of *HLA-A**02 status, is very low for many candidates, a possible association may be overlaid by outliers. 13 of the 41 cMS have an m1-mutation frequency less than 10% (see Table S4). When only examining the candidates with an m1-mutation frequency higher than 0.25 ($\overline{f}_{m1,G_{all}} > 0.25$), we observed an inverse correlation between $OLL_{G_{pr}} - OLL_{G_{non}}$ and $\overline{f}_{m1,G_{pr}} - \overline{f}_{m1,G_{non}}$ ($p = 0.01231$, Pearson's $r = -0.77254$, see labeled points Figure 3B). Selecting only *B2M*-mutated MSI CRC samples, the association is not significant, though showing the same trend ($p = 0.0691$, Pearson's $r = -0.5723$, see labeled points in Figure S5). Overall, the inverse correlation between the predicted immunogenicity of the most frequent FSPs and their mutation frequency, when comparing the two patient groups G_{pr} and G_{non}, may reflect that the mutational landscape of MSI cancer cells is partially influenced by the individual HLA constellation. In this case, cell clones with cMS mutations that are better visible to the immune system, because of an effective HLA-A-mediated antigen presentation, would occur less often. This would be reflected in the lower mutation frequency of the respective cMS. The raw data for $\overline{f}_{m1,G}$, OLL_G and the number of MSI CRC samples with mutation data for each of the 41 cMS are shown in Table S4.

4 | DISCUSSION

One central obstacle on the way to gain insights into HLA genotype-dependent occurrence and manifestation of diseases is the type of tissue material available for HLA typing. Although there is a plethora of archival FFPE samples available, commonly used assays cannot be applied because of DNA fragmentation, nucleotide changes and crosslinks, such as histone-DNA or DNA-protein crosslinks.[7,26] Robust methods that consider these alterations are necessary for making this material accessible for studying HLA type-dependent pathogenesis. As an important achievement in this issue, a Sanger sequencing-based method for determining the HLA-A type in FFPE tissue was reported.[7] We developed a simple approach for identifying *HLA-A**02 alleles in FFPE samples based on determining one described SNP specific for the *HLA-A**02 allele group.[31] Only focusing on one SNP allows a greater practicability, because ambiguous results arising from the cis/trans assignment of double peaks, one major disadvantage of Sanger sequencing, do not occur. In addition, as only one PCR reaction is required, the amount of template DNA is lower than in the method proposed by Villabona et al.[7] As a limitation of the developed approach, only one HLA-A allele group is determined in comparison to Villabona et al.[7] However, in principle it is possible to develop similar sequencing-based typing methods for identifying other HLA allele groups than *HLA-A**02 in FFPE tissue samples. Regarding the 81 (90%) samples analyzable in the validation cohort, the *HLA-A**02 status was correctly determined for all of them. The proportion of not analyzable samples is comparable to previous methods.[7]

As a next step, we evaluated the applicability of the method to current topics of immunoediting in cancer research. Recent studies reported an HLA class I genotype dependency of the mutational pattern in cancer.[12] Similar observations were made for HLA class II molecules.[14] However, as FSPs are much more immunogenic than antigens resulting from point mutations, these findings cannot be transferred to MSI cancer without further evaluation. For instance, with increasing length of an FSP, the importance of the individual HLA type for its immunogenicity may be diminished because of the plethora of possible FSP-derived peptides (for illustration see Figure S6). We divided a previously reported cohort of

73 MSI CRC patients[16] according to the presence or absence of *HLA-A**02 alleles. In addition, we defined immunological scores, accounting for the specific HLA constellation of patients with (G_{pr}) or without any (G_{non}) *HLA-A**02 allele. Based on this, we were able to contrast the frequency of cMS mutations and their estimated HLA-A-dependent immunogenicity in both patient groups G_{pr} and G_{non}. We observed an inverse association between the predicted immunogenicity of FSPs and the frequency of the underlying cMS mutations when contrasting the G_{pr} and the G_{non} group regarding the eight most frequently mutated cMS. Accordingly, the findings of Marty et al.[12,13] may also apply to MSI tumors. However, this relation was not detectable when considering all 41 candidates. This possibly illustrates that the importance of HLA-A type-dependent negative selection of cMS mutations increases with the basic mutation rate, which is determined by the length of the mononucleotide repeat.[38–40] Regarding alterations with a low basic mutation rate, the mutation itself might be the limiting factor, and not the elimination by immune cells.

In summary, we here present a method that allows identification of *HLA-A**02 alleles in FFPE tissue samples based on determining one recently described SNP in exon 2 of the *HLA-A* gene.[31] We provide one exemplary implementation in the context of MSI cancer immunoediting. However, the classification of samples according to the presence or absence of *HLA-A**02 alleles can have a broad application reaching far beyond tumor biology. By making historic and archival FFPE tissue samples accessible to molecular biological studies, the presented tool allows to gain insight into the role of HLA molecules during the pathogenesis of various illnesses, such as infectious or autoimmune diseases, and enables studying coevolution of organisms.

AUTHOR CONTRIBUTIONS

Study conception: Johannes Witt, Saskia Haupt, Aysel Ahadova, Magnus von Knebel Doeberitz, Vincent Heuveline, and Matthias Kloor. *Molecular analysis*: Johannes Witt, Lena Bohaumilitzky, Vera Fuchs, Alexej Ballhausen, Moritz Jakob Przybilla, and Michael Jendrusch. *Data analysis*: Johannes Witt, Saskia Haupt, Aysel Ahadova, Vincent Heuveline, and Matthias Kloor. *Data interpretation*: Johannes Witt, Saskia Haupt, Aysel Ahadova, Magnus von Knebel Doeberitz, Vincent Heuveline, and Matthias Kloor. *Manuscript writing*: Johannes Witt, Saskia Haupt, Aysel Ahadova, Magnus von Knebel Doeberitz, Vincent Heuveline, and Matthias Kloor. *Resources*: Toni T. Seppälä, Daniel Fürst, Thomas Walle, Elena Busch, Georg Martin Haag, Robert Hüneburg, Jacob Nattermann, Magnus von Knebel Doeberitz, Vincent Heuveline, and Matthias Kloor. *Revision and final approval of the manuscript*: All authors.

ACKNOWLEDGMENTS

This work was supported by Else-Kröner-Fresenius Foundation (2018_A44), the German Cancer Aid (70113455), and the Klaus Tschira Foundation (00.012.2021), Heidelberg, Germany, in frame of the project "Mathematics in Oncology - Towards optimal prevention and treatment in patients with inherited cancer syndrome". TTS was funded by Sigrid Juselius Foundation, Jane and Aatos Erkko Foundation, Finnish Cancer Society, Relander Foundation, Emil Aaltonen Foundation, Finnish Medical Foundation, State Research Funding (TYH2022323), Academy of Finland (33867), and the iCAN Precision Medicine Flagship. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

Georg Martin Haag (no relationship to the current manuscript). Consulting or Advisory Role: Bristol-Myers Squibb; MSD Sharp & Dohme; Lilly; Novartis. Honoraria: Servier; MSD Sharp & Dohme; Lilly; Targos; Bristol-Myers Squibb; IOMEDICO, MCI Conventions. Research Funding: Nordic Pharma; Taiho Pharmaceutical; MSD Sharp & Dohme; Janssen; Astra Zeneca; IKF Klinische Krebsforschung Frankfurt. Travel; Accommodations: Bristol-Myers Squibb; Lilly; Servier; MSD Sharp & Dohme. Robert Hüneburg (no relationship to the current manuscript). Consulting or Advisory Role: Cancer Prevention Pharmaceuticals; Janssen Pharmaceuticals. Endoscopic equipment on loan from Fujifilm Germany. Jacob Nattermann (no relationship to the current manuscript). Endoscopic equipment on loan from Fujifilm Germany. All other authors confirm that there are no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article.

ORCID

Johannes Witt https://orcid.org/0000-0001-6648-8828
Moritz Jakob Przybilla https://orcid.org/0000-0001-5645-9492
Daniel Fürst https://orcid.org/0000-0001-9542-3169
Thomas Walle https://orcid.org/0000-0003-4835-955X

REFERENCES

1. Klein J, Sato A. The HLA system. First of two parts. *N Engl J Med*. 2000;343(10):702-709.
2. Madden K, Chabot-Richards D. HLA testing in the molecular diagnostic laboratory. *Virchows Arch*. 2019;474(2):139-147.

3. Alcoceba M, Sebastian E, Marin L, et al. HLA specificities are related to development and prognosis of diffuse large B-cell lymphoma. *Blood*. 2013;122(8):1448-1454.
4. Bonamigo RR, Carvalho AV, Sebastiani VR, Silva CM, Pinto AC. HLA and skin cancer. *An Bras Dermatol*. 2012;87(1): 9-16. quiz 17–18.
5. Aureli A, Canossi A, Del Beato T, et al. Breast cancer is associated with increased *HLA-DRB1**11:01 and *HLA-DRB1**10:01 allele frequency in a population of patients from Central Italy. *Immunol Invest*. 2020;49(5):489-497.
6. Aureli A, Canossi A, Del Beato T, et al. *HLA-DRB1**13:01 allele in the genetic susceptibility to colorectal carcinoma. *Int J Cancer*. 2015;136(10):2464-2468.
7. Villabona L, Leon Rodriguez DA, Andersson EK, Seliger B, Dalianis T, Masucci GV. A novel approach for HLA-A typing in formalin-fixed paraffin-embedded-derived DNA. *Mod Pathol*. 2014;27(9):1296-1305.
8. Falk K, Rotzschke O, Stevanovic S, Jung G, Rammensee HG. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature*. 1991;351(6324):290-296.
9. Sidney J, Southwood S, Sette A. Classification of A1- and A24-supertype molecules by analysis of their MHC-peptide binding repertoires. *Immunogenetics*. 2005;57(6):393-408.
10. Sidney J, Peters B, Frahm N, Brander C, Sette A. HLA class I supertypes: a revised and updated classification. *BMC Immunol*. 2008;9:1.
11. Blackwell JM, Jamieson SE, Burgner D. HLA and infectious diseases. *Clin Microbiol Rev*. 2009;22(2):370-385.
12. Marty R, Kaabinejadian S, Rossell D, et al. MHC-I genotype restricts the oncogenic mutational landscape. *Cell*. 2017;171(6): 1272-1283 e1215.
13. Marty R, de Prisco N, Carter H, Font-Burgada J. MHC-I genotype drives early immune selection of oncogenic mutations. *Mol Cell Oncol*. 2018;5(2):e1409863.
14. Marty Pyke R, Thompson WK, Salem RM, Font-Burgada J, Zanetti M, Carter H. Evolutionary pressure against MHC class II binding cancer mutations. *Cell*. 2018;175(2):416-428 e413.
15. Naranbhai V, Viard M, Dean M, et al. *HLA-A**03 and response to immune checkpoint blockade in cancer: an epidemiological biomarker study. *Lancet Oncol*. 2022;23(1):172-184.
16. Ballhausen A, Przybilla MJ, Jendrusch M, et al. The shared frameshift mutation landscape of microsatellite-unstable cancers suggests immunoediting during tumor evolution. *Nat Commun*. 2020;11(1):4740.
17. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. *Nature*. 2013;500(7463):415-421.
18. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012;487(7407):330-337.
19. Kloor M, von Knebel DM. The immune biology of microsatellite-unstable cancer. *Trends Cancer*. 2016;2(3):121-133.
20. Alhopuro P, Sammalkorpi H, Niittymaki I, et al. Candidate driver genes in microsatellite-unstable colorectal cancer. *Int J Cancer*. 2012;130(7):1558-1566.
21. Linnebacher M, Gebert J, Rudy W, et al. Frameshift peptide-derived T-cell epitopes: a source of novel tumor-specific antigens. *Int J Cancer*. 2001;93(1):6-11.
22. Townsend A, Ohlen C, Rogers M, Edwards J, Mukherjee S, Bastin J. Source of unique tumour antigens. *Nature*. 1994; 371(6499):662.
23. Buckowitz A, Knaebel HP, Benner A, et al. Microsatellite instability in colorectal cancer is associated with local lymphocyte infiltration and low frequency of distant metastases. *Br J Cancer*. 2005;92(9):1746-1753.
24. Phillips SM, Banerjea A, Feakins R, Li SR, Bustin SA, Dorudi S. Tumour-infiltrating lymphocytes in colorectal cancer with microsatellite instability are activated and cytotoxic. *Br J Surg*. 2004;91(4):469-475.
25. Kloor M, Becker C, Benner A, et al. Immunoselective pressure and human leukocyte antigen class I antigen machinery defects in microsatellite unstable colorectal cancers. *Cancer Res*. 2005; 65(14):6418-6424.
26. Do H, Dobrovic A. Sequence artifacts in DNA from formalin-fixed tissues: causes and strategies for minimization. *Clin Chem*. 2015;61(1):64-71.
27. Menon AG, Morreau H, Tollenaar RA, et al. Down-regulation of HLA-A expression correlates with a better prognosis in colorectal cancer patients. *Lab Invest*. 2002;82(12):1725-1733.
28. Gamzatova Z, Villabona L, Dahlgren L, et al. Human leucocyte antigen (HLA) A2 as a negative clinical prognostic factor in patients with advanced ovarian cancer. *Gynecol Oncol*. 2006;103(1):145-150.
29. De Petris L, Bergfeldt K, Hising C, et al. Correlation between HLA-A2 gene frequency, latitude, ovarian and prostate cancer mortality rates. *Med Oncol*. 2004;21(1):49-52.
30. Gonzalez-Galarza FF, McCabe A, Santos E, et al. Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Res*. 2020;48(D1):D783-D788.
31. Song S, Han M, Zhang H, Wang Y, Jiang H. Full screening and accurate subtyping of *HLA-A**02 alleles through group-specific amplification and mono-allelic sequencing. *Cell Mol Immunol*. 2013;10(6):490-496.
32. Robinson J, Barker DJ, Georgiou X, Cooper MA, Flicek P, Marsh SGE. IPD-IMGT/HLA Database. *Nucleic Acids Res*. 2020;48(D1):D948-D955.
33. Cock PJ, Antao T, Chang JT, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*. 2009;25(11):1422-1423.
34. Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res*. 2020;48(W1):W449-W454.
35. Stadhouders R, Pas SD, Anber J, Voermans J, Mes TH, Schutten M. The effect of primer-template mismatches on the detection and quantification of nucleic acids using the 5′ nuclease assay. *J Mol Diagn*. 2010;12(1):109-117.
36. Rejali NA, Moric E, Wittwer CT. The effect of single mismatches on primer extension. *Clin Chem*. 2018;64(5):801-809.
37. Kherreh N, Cleary S, Seoighe C. No evidence that HLA genotype influences the driver mutations that occur in cancer patients. *Cancer Immunol Immunother*. 2022;71(4):819-827.
38. Duval A, Rolland S, Compoint A, et al. Evolution of instability at coding and non-coding repeat sequences in human MSI-H colorectal cancers. *Hum Mol Genet*. 2001;10(5):513-518.
39. Jonchere V, Marisa L, Greene M, et al. Identification of positively and negatively selected driver gene mutations associated with colorectal cancer with microsatellite instability. *Cell Mol Gastroenterol Hepatol*. 2018;6(3):277-300.

40. Woerner SM, Yuan YP, Benner A, Korff S, von Knebel DM, Bork P. SelTarbase, a database of human mononucleotide-microsatellite mutations and their potential impact to tumorigenesis and immunology. *Nucleic Acids Res*. 2010;38(Database issue):D682-D689.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Witt J, Haupt S, Ahadova A, et al. A simple approach for detecting *HLA-A**02 alleles in archival formalin-fixed paraffin-embedded tissue samples and an application example for studying cancer immunoediting. *HLA*. 2023;101(1):24-33. doi:10.1111/tan.14846